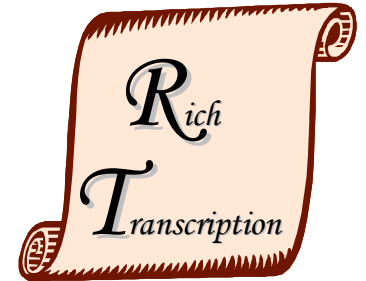


The Rich Transcription 2009 Evaluation Overview



<http://www.nist.gov/speech/tests/rt/rt2009>

2009 Rich Transcription Evaluation Conference

May 28-29, 2009

Melbourne, FL

Jérôme Ajot & Jonathan Fiscus

<http://itl.nist.gov/iad/mig/tests/rt/2009/>

Rich Transcription Conceptualization (Broadcast News Domain)

Traditional Speech-To-Text Output

from a. b. c. news world headquarters in new york this is world news tonight with peter jennings good evening every one we're going to begin tonight on the presidential campaign trail littered as it is today with democratic candidates who have suddenly been obliged to reassess their futures when former vice president al gore formally endorsed howard dean as the presidential candidate today

Enriched Output

<speaker name="Announcer"> <SU type="stmt"> from <prop_noun> a. b. c. news world headquarters </prop_noun> in <prop_noun> new york </prop_noun> <phrase_bound> this is <prop_noun> world news tonight </prop_noun> with <prop_noun> peter jennings </prop_noun> </phrase_bound> </SU> </speaker>

<speaker name="Peter Jennings"> <SU type="stmt"> good evening every one </SU> <SU type="stmt"> we're going to begin tonight on the presidential campaign trail

Readable Transcript

Announcer: From ABC News World Headquarters in New York, this is World News Tonight with Peter Jennings.

Peter Jennings: Good evening every one. We're going to begin tonight on the presidential campaign trail. Littered as it is today with Democratic candidates who have suddenly been obliged to reassess their futures when former Vice President Al Gore formally endorsed Howard Dean as the presidential candidate today.

Other language processing

Translate
Summarize
Parse
Extract Info
Search

Rich Transcription Conceptualization (Meeting Domain)

Speaker ID	Speech Recognition	Video OCR	Action Detection
Speaker ID=5	we can write them down and be able to ...	[Camera 2] - DATE OF SHOWER AUG 1st or 20th -LOCATION - PCC CENTER ?? SUNRISE ? -CATERER -? -CAKE CUPCAKES - RACH LEAH CHOCOLATE VANILLA (MATIL) DECORATIONS LEAH RACH FAVORS	Participant 5 Speaks
Speaker ID=4	mm hmm		Participant 4 Speaks Participant 5 looks at 3
Speaker ID=3	the pay ... and availability of course		Participant 3 Speaks Participant 2 looks at 4
Speaker ID=4	mm hmm		Participant 4 Speaks
Speaker ID=1	I'll write it down		Participant 1 Speaks Participant 1 Stands Participant 1 Walks
Speaker ID=2	And the invitations are you, the cake is you both, the music will be whoever you pick but you will pick the music though.	[Camera 2] DJ - JOHN	Participant 2 Speaks Participant 2 Points Participant 2 Points
Speaker ID=2	Let Antoine pick the music really I think that's probably good. He's very p- you know he probably wants the right kind. And clean-up ...		Participant 1 Writes



Rich Transcription Conceptualization (Meeting Domain)

Speaker ID	Speech Recognition	Video OCR	Action Detection
Speaker ID=5	we can write them down and be able to ...	[Camera 2] - DATE OF SHOWER AUG 1st or 20th -LOCATION - PCC CENTER ?? SUNRISE ? -CATERER -? -CAKE CUPCAKES - RACH LEAH CHOCOLATE VANILLA (MATIL) DECORATIONS LEAH RACH FAVORS	Participant 5 Speaks
Speaker ID=4	mm hmm		Participant 4 Speaks Participant 5 looks at 3
Speaker ID=3	the pay ... and availability of course		Participant 3 Speaks Participant 2 looks at 4
Speaker ID=4	mm hmm		Participant 4 Speaks
Speaker ID=1	I'll write it down	[Camera 2] DJ - JOHN	Participant 1 Speaks Participant 1 Stands Participant 1 Walks
Speaker ID=2	And the invitations are you, the cake is you both, the music will be whoever you pick but you will pick the music though.		Participant 2 Speaks Participant 2 Points Participant 2 Points
Speaker ID=2	Let Antoine pick the music really I think that's probably good. He's very p- you know he probably wants the right kind. And clean-up ...		Participant 1 Writes



Rich Transcription Evaluation Series

- Goal
 - Develop recognition technologies that produce language content representations (transcripts) which are understandable by humans and useful for downstream processes.
- Domains
 - Meeting Room speech
 - Broadcast News (BN)
 - Conversational Telephone Speech (CTS)
- Parameterized “Black Box” evaluations
 - Evaluations control input conditions to investigate weaknesses/strengths
 - Sub-test scoring provides finer-grained diagnostics

RT-09 Evaluation Tasks

- Evaluation of essential component technologies
 - Speech-To-Text Transcription (STT)
 - Transcribe the spoken words
 - Diarization “Who Spoke When” (SPKR)
 - Detect segments of speech and cluster them by speaker
 - Speaker Attributed Speech-To-Text (SASTT)
 - Transcribe the spoken words and associate them with a speaker
- NEW for RT-09
 - No Lecture Room/Coffee break data
 - Video input conditions

RT-09 Evaluation Participants

Site ID	Site Name	Evaluation Task			
		SPKR		STT	SASTT
		Audio	Audio/Video		
AMI	Augmented Multi-party Interaction: Univ. Sheffield, IDIAP, Univ. Edinburgh, Univ. of Technology Brno, Univ. Twente	X		X	X
I2R/NTU	Infocomm Research Site and Nanyang Technological University	X			
FIT	Florida Institute of Technology			X	
ICSI	International Computer Science Institute	X	X		
LIA/Eurecom	Laboratoire Informatique d'Avignon/ Ecole d'ingénieurs et centre de recherche en Systèmes de Communications	X			
SRI/ICSI	SRI International and International Computer Science Institute			X	X
UPM	Universidad Politécnica de Madrid	X			
UPC	Universitat Politècnica de Catalunya	X			

System Input Conditions

Audio

Distant Mic. Conditions

- Mult. Distant Mics. (MDM)
- Mult. Mark III Arrays (MM3A)
- All Distant Mics. (ADM)

Contrast Conditions

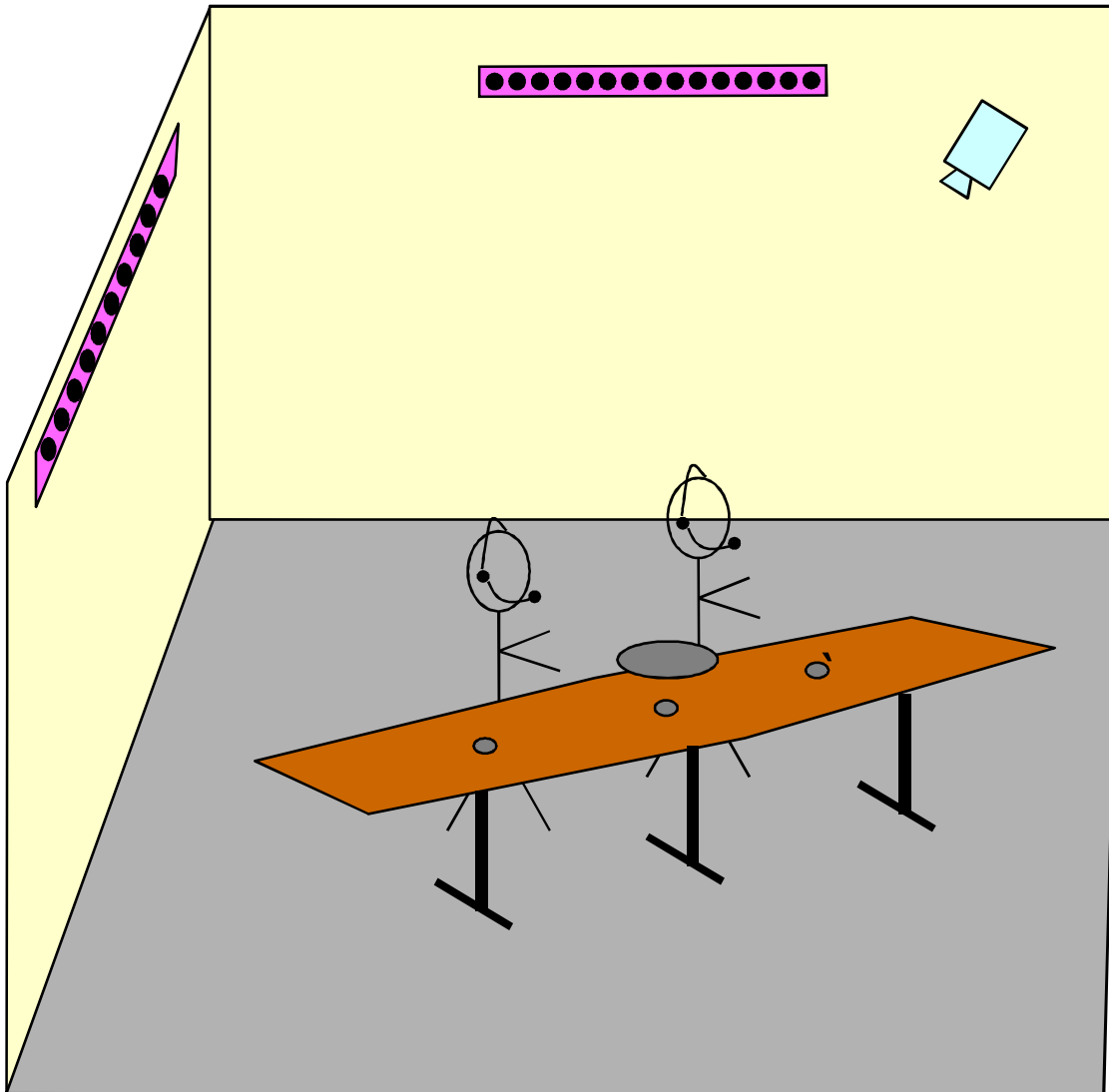
- Individual Head Mics. (IHM)
- Single Distant Mic. (SDM)

Laboratory Conditions

- IHM with reference segmentations (IHM-REFSEG)

Video Inputs

- All Video (AV)



Six System Input Conditions

- Audio input conditions
 - Distant microphone conditions
 - Multiple Distant Microphones (MDM)
 - Three or more centrally located table mics
 - Multiple Mark III digital microphone Arrays (MM3A)
 - Linear topology, 64-channel digital microphone array
 - All Distant Microphones (ADM)
 - Contrastive microphone conditions
 - Single Distant Microphone (SDM)
 - Center-most MDM microphone
 - Gauge performance benefit using multiple table mics
 - Individual Head Microphones (IHM)
 - Performance on clean speech
 - Similar to Conversational Telephone Speech
 - » One speaker per channel, conversational speech
 - Laboratory conditions
 - Individual Head Mics. With reference segmentations (IHM-REFSEG)
- Video input conditions
 - All video (AV)

Training/Development Corpora

- Meeting resources
 - ICSI Meeting Corpus
 - ISL Meeting Corpus
 - NIST Meeting Pilot Corpus (including RT-02 test data)
 - NIST Phase II Meeting Corpus
 - CHIL '05, '06, and '07 development test sets
 - AMI Meeting Corpora
 - Rich Transcription 2004 Spring (RT-04S) Development & Evaluation Data
 - Rich Transcription 2005, 2006, 2007 Evaluation Data
- Others
 - Topic Detection and Tracking Phase 4 (TDT4) corpus
 - Fisher English conversational telephone speech corpus

RT-09 Conference Meeting Test Set



Edinburgh – 60 minutes

(All non-American English)

(1 Remote Participant)

1. Remote Control Design
“B” meeting
4 Participants
30 minutes
2. Remote Control Design
“C” meeting
4 Participants
30 minutes

IDIAP – 60 minutes

(All non-American English)

1. Discussion: Poverty
4 Participants
30 minutes
2. General discussion
4 Participants
30 minutes

NIST – 60 minutes

(All American English)

1. Baby shower planning
5 Participants
20 minutes
2. Focus group discussion
6 Participants
19 minutes
3. Focus group discussion
11 Participants
21 minutes

Thanks to All

- Data Contributors:
 - Edinburgh, IDIAP, and NIST
- Test Data Preparation:
 - LDC for transcription of the evaluation data
 - LIMSI for their forced alignment tool
- Venue arrangements:
 - Veton Kepuska and the Florida Institute of Technology staff